PRE-PRINT Stramandinoli F., Cangelosi A., Marocco D. (2011). Towards the grounding of abstract words: A neural network model for cognitive robots. Proceedings of IJCNN-2011 International Joint Conference on Neural Networks, San Jose. IEEE

Towards the Grounding of Abstract Words: A Neural Network Model for Cognitive Robots

Francesca Stramandinoli, Angelo Cangelosi and Davide Marocco

***Abstract*— In this paper, a model based on Artificial Neural Networks (ANNs) extends the symbol grounding mechanism to abstract words for cognitive robots. The aim of this work is to obtain a semantic representation of abstract concepts through the grounding in sensorimotor experiences for a humanoid robotic platform. Simulation experiments have been developed on a software environment for the iCub robot. Words that express general actions with a sensorimotor component are first taught to the simulated robot. During the training stage the robot first learns to perform a set of basic action primitives through the mechanism of direct grounding. Subsequently, the grounding of action primitives, acquired via direct sensorimotor experience, is transferred to higher-order words via linguistic descriptions. The idea is that by combining words grounded in sensorimotor experience the simulated robot can acquire more abstract concepts. The experiments aim to teach the robot the meaning of abstract words by making it experience sensorimotor actions. The iCub humanoid robot will be used for testing experiments on a real robotic architecture.**

I. Introduction

COGNITIVE DEVELOPMENTAL ROBOTICS, taking inspiration from developmental mechanisms studied in children by psychologists and neuroscientists, focuses on the development of cognitive processes in humanoid robots. This is an innovative approach to robotics that presents a strong interdisciplinary character and aims to overcome current limitations in robot design. In most of the literatures so far, cognitive processes have been mainly investigated in the context of separate research areas. However, recent studies have been shown that mental processes are deeply influenced by the structure of the body and its interaction with the environment (embodiment) [1][2][3]. These new findings are now altering the interaction between different disciplines ranging from neuroscience to robotics. For this reason, humanoid robots are achieving more importance in scientific research on cognitive science and they are becoming a crucial tool in the study of human behaviour [4].

Cognitive robots have been successfully used for learning concrete concepts. In [5][6], it has been shown that cognitive robots are capable of performing concrete actions and understanding each action's name. Building intelligent systems that can learn the meaning of abstract words is a challenging task for cognitive developmental robotics. Abstract concepts represent everything that is not physically defined and/or spatially constrained (e.g. mental states) and refer to things that are intangible, which are perceived not through the senses but by the mind (e.g. "truth", "democracy", "happiness", "justice"). This is why understanding the representation of abstract words is a highly challenging and problematic issue in cognitive and neural sciences, and in cognitive robotics.

In this paper a neural network model for the learning of abstract words in cognitive robots is presented. The aim of the robotics experiments is to find a semantic representation of abstract words for a humanoid robotic platform. Simulation experiments have been developed on a software environment for the iCub robot [4]. Words like "ACCEPT", "REJECT", "PICK", that express general actions with a sensorimotor component, are taught the robot. The iCub first learns a set of basic action primitives through the mechanism of direct grounding; then, the grounding is transferred from basic symbols to new ones, the latter obtained by logical combinations of the elementary words. Specifically, at the beginning of the training the simulated robot learns to perform a series of action primitives (e.g. "MOVE_ARM_AWAY", "MOVE_ARM_TOWARD", "OPEN_HAND", "CLOSE_HAND", etc.) and then, through the process of symbol grounding transfer, by combining action primitives, the robot acquires more abstract concepts (e.g. "KEEP", "GIVE", "RECEIVE"). The goal of this study is to prove that the grounding of abstract categories can be obtained as a consequence of combining the grounding of sensorimotor experiences.

The paper is organised as follows. Section II identifies the scientific and technological issues that the paper intends to address and solve; it also gives an overview of the most relevant classical theories in this research field. A description of the model from which the experiments take inspiration, the neural network architecture implemented and the robotic platform used for experiments, are described in Section III. Section IV presents the experimental results and the discussion. Conclusions of Section V close the paper.

II. The Symbol Grounding Problem

The symbol grounding problem, formulated by Harnard in 1990 [7], is related to the matters of "how words get their meanings", "how words are connected to the things they refer to". The problem, as Harnard said, is analogous to try to learn Chinese from a Chinese/Chinese dictionary alone [7]. This problem affects all symbolic cognitivist models that deal with language learning and evolution. Different approaches have been proposed in order to solve the symbol grounding

Francesca Stramandinoli, Angelo Cangelosi and Davide Marocco, School of Computing and Mathematics, University of Plymouth, United Kingdom (email: {francesca.stramandinoli, acangelosi, davide.marocco}@plymouth.ac.uk).

This research has been supported by the EU project RobotDoC number 235065 from the 7th Framework Programme, Marie Curie Action ITN.

problem [8]; the two main modelling approaches to this problem are Symbolic models and Connectionist models.

Symbolic theories assume that conceptual representations are typically non perceptual [9]. According to symbolic models, concepts are generated by abstract, arbitrary and amodal symbols [10] and the mind is viewed as a symbol system and cognition is symbol manipulation [7]. In these models there are a set of arbitrary symbols that are combined and manipulated on the basis of specific rules, neglecting of the interaction with real world. One of the most prominent symbolic computational model of language processing is the Latent Semantic Analysis (LSA) proposed by Landauer and Dumais [10]. This is a corpus-based statistical method for representing the meaning of words. This method analyses the relationships between a set of documents and the terms they contain by producing a set of concepts related to documents and terms. In order to show that computationalism is incorrect, Searle in 1980 formulated his celebrated Chinese Room Argument [11]. According to the symbolic theory of mind, if a computer could pass the Turing Test in Chinese, then this would mean that the computer understands the meaning of Chinese symbols in the same sense that an English-speaking person understands the meaning of English symbols [7]. Searle attempted to show that a symbol-processing machine can never be properly described as "having a mind" or "understanding", regardless of how intelligently it may behave.

The Connectionist approach based on artificial neural networks, on the contrary of symbolic models, focuses on the grounding of symbols into perception [12] and it models mental phenomena as an emergent process of interconnected networks [13]. Roughly speaking, in the connectionist model learning the meaning of a word is a matter of establishing a connection between a set of stimuli and a verbal response. The meaning of words consists of associating verbal labels with their experience at the time that the label is used.

According to the Connectionist Embodied models, based on the combination of neural networks and robotics methodologies, cognition is emergent from the sensorimotor interaction of an agent with the environment. This model allows us to overcome the limitations of symbolic and connectionist models in the development of language learning systems. Some accounts of the classical embodied theories of abstract words proposed in literature are Cognitive Simulation Theories [1], Memory Theories [2] and Cognitive Linguistics Theories [3].

In his Perceptual Symbol Systems (PSS) theory [1], Barsalou emphasises the role of simulation in cognition. According to this theory, perceptual symbols are multimodal, sensorimotor, proprioceptive, and introspective, that is they activate different motor and sensorial information (e.g. vision, audition, touch, etc) tightly linked to the interaction with the world [14]. Perceptual symbols are neural representations located in sensory-motor areas in the brain of embodied actions and they are activated during the experience with the world. When the body interacts with the environment (e.g. *sitting down on a chair*), the brain captures and stores in memory neural activation patterns present during the experience with objects and entities (e.g. *how a chair looks*, *the action of sitting*). These perceptual symbols then combine, when they are semantically related, to form a simulator (concept). Later, when knowledge is needed to represent a category (e.g. *chair*), these neural activation patterns are reactivated to simulate the concept [15]. In [16] Andrews et al. propose a Bayesian model for representing words. In this work, the authors identify two statistical data types from which a semantic representation of words can be learned. In particular, they claim that a semantic representation of words can be derived from an optimal statistical combination of experiential data and distributional data. Experiential data are sensorimotor and they are collected through the interaction of the body with the physical world; on the contrary, distributional data describe the statistical distribution of words in language. In this framework, experiential and distributional data are both nontrivial source of semantic information for obtaining a semantic representation of words.

In the Memory Theories of grounded cognition, situated action plays an important role. According to embodiment, cognition is not only grounded in our body, but it is also "situated" (role of the body and of the context in representations). As proposed by Glenberg [2] the meaning of a situation depends on a set of stimuli available for acting on objects (affordances) tuned on the individuals personal experience and according to the goal to be pursued. For example, if the goal of a human beings is to change a light bulb, the meaning of the situation will arise from affordances related to a light bulb (e.g. *holding it in the hand*) "meshed" with the affordances of a chair (e.g. *it supports for reaching the bulb*) related to the goal to be pursued.

Within the framework of Cognitive Linguistics Theories [17], Lakoff and Johnson [3] argue that abstract concepts are grounded metaphorically in embodied and situated knowledge. According to their theory, human beings acquire an extensive knowledge about their bodies (e.g. eating) and situations (e.g. verticality), by interacting with the environment. Abstract concepts can be obtained from this knowledge metaphorically. For example, *love* can be understood as eating (e.g "being consumed by a lover"), and affective experience can be understood as verticality (e.g. "happy is up, sad is down") [15].

Connectionist neural network models have been demonstrated to be good candidates to deal with the symbol grounding problem. In [18] authors investigated the transfer of grounding from names of geometric shapes (e.g., circles, ellipses, squares and rectangles) to the superposed categories ("symmetric" and "asymmetric"). This model was then expanded in [19] to deal with larger category sets, and to look at different aspects of the transfer of grounding. In [20] connectionist simulations of the grounding transfer mechanism were performed. The simulation results demonstrated that grounding is transferred from symbols denoting object properties to new ones denoting the object as a whole.

More recently, neural network models have been successfully used for the acquisition of language in cognitive robots. In [6] a neural network controller for the learning of actions and their corresponding name has been presented. In the proposed system, a simulated agent is controlled by a three-layer feedforward neural network that has vision, motor and language inputs for linguistic comprehension and production. The use of a dual-route neural architecture permits the simultaneous simulation of language production (from vision to language) and language understanding (from language to vision/action) abilities. In the experiments that are going to be presented a connectionist and embodied approach for the learning of abstract words will be adopted.

III. Model Description

A. Action and Language Repertoires and Training procedure

According to the connectionist and embodied model, linguistic abilities are developed through the direct interaction between cognitive agents and the physical world they interact with. The aim of the experiments reported in this paper is to find a mechanism for the learning of abstract words through sensorimotor experiences, using the humanoid robotic platform iCub [23]. Experiments take inspiration from the model proposed by Cangelosi and Riga [5] in which two simulated robots, teacher and learner, were trained to learn a set of basic action primitives. In the proposed model, the training of the learner requires two mechanisms. The first is the direct grounding of basic words, during which the agent, by observing the teacher, learns a set of basic action primitives and their corresponding name via direct sensorimotor experience. The second mechanism is the grounding transfer process by which the grounding of basic words is transferred to higher-order words via linguistic description [21]. In particular, the training of the robot consists of three incremental stages:

- (i) Basic Grounding (BG)
- (ii) Higher-order Grounding 1 (HG1)
- (iii) Higher-order Grounding 2 (HG2)

During the BG stage the robot learns to perform, by imitation, basic action primitives and their corresponding names (e.g. "CLOSE_LEFT_ARM", "CLOSE_RIGHT_ARM", "MOVE_FORWARD"). In the HG1 stage the robot, via linguistic description, acquires new words combining basic action primitives (e.g. "GRAB" [is] "CLOSE_LEFT_ARM" [and] "CLOSE_RIGHT_ARM"). During the HG2 stage the robot learns high level words through the combination of action primitives and higher-order action words (e.g. "CARRY" [is] "GRAB" [and] "MOVE_FORWARD").

In the experiments that are going to be presented, the robot will first learn to perform a series of action primitives (e.g. "PUSH", "PULL", "GRASP", "RELEASE") through the mechanism of direct grounding. Subsequently, by correlating higher-order action words (e.g. "KEEP", "GIVE", "RECEIVE") with basic action primitives, the robot will acquire more abstract concepts (e.g. "PICK", "ACCEPT", "REJECT"). A set of objects, whose visual properties can be sensed, are being used as training set of the experiment. Each object of the training set is labeled to a certain degree of likeness, according to its characteristics. The idea is that, the robot by interacting with the objects in his environment can acquire action primitives such as "GRASP", "RELEASE", etc. After, by combining these elementary words the robot is capable of learning new words like "KEEP", "GIVE", "RECEIVE". For example, when a new object is presented to the robot, it evaluates its properties (e.g. weight, shape, colour, etc) and makes a decision to "ACCEPT" it or not. At the end of the experiment, the robot learns how to behave through interaction with the environment and is able to categorise abstract symbols by experiencing sensorimotor actions.

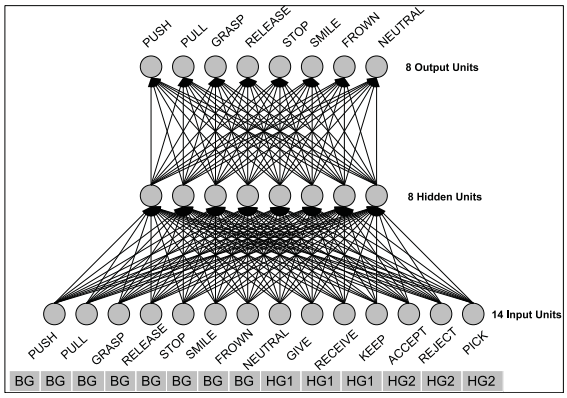


Fig. 1. Architecture of association network for the learning of verbal labels associated to action primitives

B. Neural Network Architecture

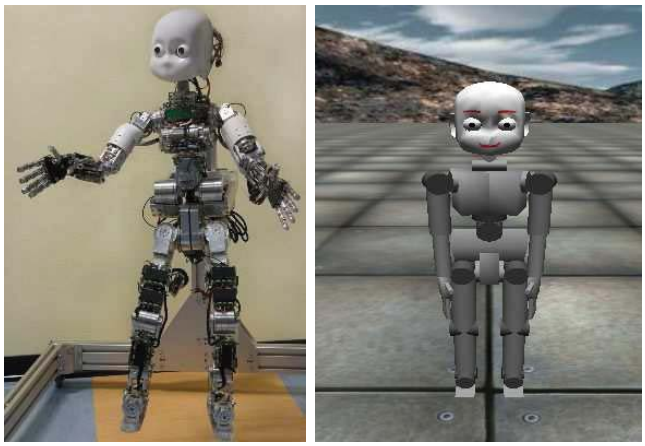
We propose an Artificial Neural Network (ANN) model that extends the symbol grounding mechanism to abstract words for the iCub robot. As discussed above, embodied ANNs are particularly suitable for modelling the relationship between language (perceptual symbolic input) and action (sensorimotor output e.g. the concept/action of pushing, pulling, grasping) [21][22]. A neural network controller has been implemented for the learning of verbal labels associated to action primitives (Fig. 1).

The robot's neural network, as shown in Fig. 1, is a 3-layers feedforward NN, fully connected, with a sigmoid activation function (1)

$$f(x) = \frac{1}{(1 + e^{-\lambda x})} \tag{1}$$

with unity slope ($\lambda = 1$).

The network has 14 input units that encode the name of all the actions that can be taught the robot (localist encoding of words). The hidden units consist of 8 neurons that are fully connected with both input and output units. The output of the network selects which action primitive needs to be activated in order to obtain a specific behaviour. When one of the input units is activated, the network selects as output a combination of action primitives to be executed.



(a) (b)

Fig. 2. The iCub: real robotic architecture (a) and software simulator (b)

The output of the network is the input for the iCub module that implements the execution of action primitives.

C. The iCub Robotic Platform

The iCub humanoid robot, shown in Fig. 2(a), has been used as robotic platform for the experiments performed. The iCub, developed at the end of the RobotCub European project, is an open-source robotic platform for research in embodied cognition [23]. The robot is $104\,cm$ tall and his overall weight is $22\,kg$; the dimensions of it are similar to that of a three and half year old child.

The iCub has 53 Degrees Of Freedom (DOF) allocated as follow: 6 DOF for the head, 3 DOF for the torso, 12 DOF for the legs and 32 DOF for the upper limbs. In particular, each upper limb has 16 DOF, 7 of which are in the arm and 9 in the hand (3 for the thumb, 2 for the index, 2 for the middle finger, 1 for the coupled ring and little finger, 1 for the adduction/abduction). The iCub is also equipped with different types of sensors; in particular, it has digital cameras and microphones from which it can see and hear, gyroscopes and accelerometers for the movement and force/torque sensors. A distributed sensorized skin is under development.

The iCub Simulator (Fig. 2(b)) has been used as a tool for testing algorithms before to use the real robotic architecture [24][25]; the simulator has been developed by using open source libraries and it reproduces the physics and the dynamics of the real robot. The software simulator is a replication of the first iCub prototype, and it was developed collecting data directly form the robot design specifications.

IV. Simulation Results and Discussion

In order to verify the validity of the model proposed for the learning of the meaning of abstract words in the iCub humanoid robot, simulations experiments have been performed. The robot's ANN controller was developed in C++ programming language and linked with the iCub simulator. The execution of basic actions was implemented by using the Action Primitives library from the iCub software repository; this library provides a set of action primitives that can be easily combined in order to obtain more complex behaviour. In our model the grounding of words is obtained by linking verbal labels to sensorimotor experience.

A. Neural Network Training

For the training of the neural network model a supervised learning algorithm was chosen. In particular, the network is trained by using the error backpropagation algorithm to teach the neural controller to perform a given task through the optimization of the error with respect to the weights. The neural controller has been trained with the following configuration: weights of the network are initialized to a random value in the range $[\pm 0.5]$ and the backpropagation algorithm is applied for 180000 iterations, with a learning rate (α) equals to 0.2 and momentum (β) 0.9. The learning rate and momentum, in general, can assume values between $[0, 1]$; in our simulation, a small value of α slows the convergence rate of the algorithm but helps to ensure that the global minimum is not missed. A small learning rate value has been coupled with a larger momentum to control the convergence rate of the algorithm; a big value of β increases convergence speed.

As described in the previous section, the proposed algorithm for the association of verbal labels to actions primitives consists of three incremental steps: (i) the Basic Grounding, (ii) the Higher-order Grounding 1 and (iii) the Higher-order Grounding 2.

During the BG training stage, the robot learns the names associated to the eight action primitives through direct sensorimotor experience. The names of action primitives, that are the input of the neural network, are "PUSH", "PULL", "GRASP", "RELEASE", "STOP", "SMILE", "FROWN", "NEUTRAL". The output of the network in this stage is computed by applying the backpropagation algorithm online, to calculate the weights correction from the input. In the post-training test, the results of the simulations show that the network performs correctly the mapping between inputs and outputs. The BG learning stage runs for 30000 iterations. As shown in Fig. 3, after 5000 runs, the value of the error is already smaller then 0.02.

The HG1 and HG2 training phases implement the grounding transfer process. During these stages the grounding of basic words, acquired via direct sensorimotor experience, is transferred to higher-order words via linguistic description that, in the neural controller implementation, is semplified as an input pattern to the network. The grounding transfer consists of multiple steps, depending on the number of action primitives that are combined to obtain a more complex behaviour. For example, in order to transfer the grounding from the basic actions "GRASP" and "STOP" to the higher-order word "KEEP" ("KEEP" [is] "GRASP" [and] "STOP") two steps are required, one for each basic action primitive involved. Each of this steps consist of two phases (Fig. 4):

- the network receives as input the action primitives words contained in the linguistic description of the higher-order word and computes the corresponding output

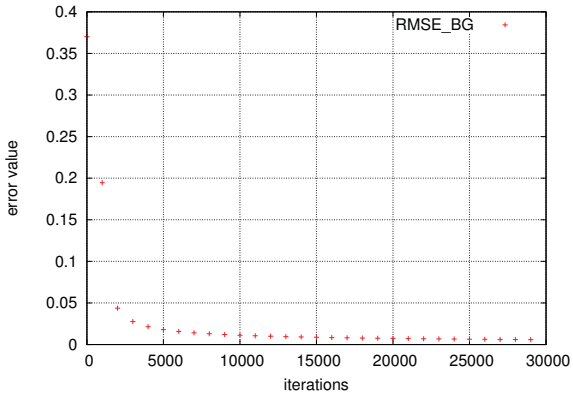


Fig. 3. Root Mean Square Error after the BG training stage

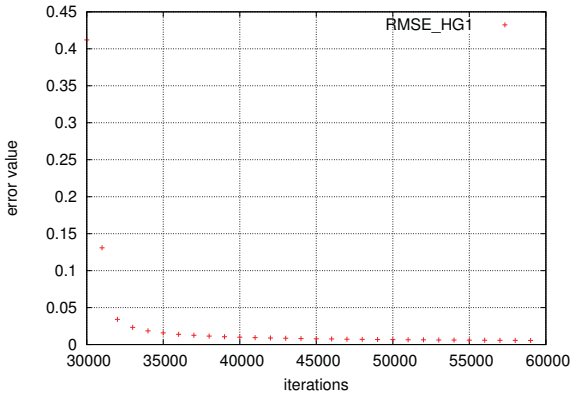


Fig. 5. Root Mean Square Error after the HG1 training stage

without applying backpropagation algorithm (feedforward phase without learning).

- the network receives as input the name of the higher-order word and as target the output of the network calculated during the feedforward phase (backpropagation learning).

This procedure is adopted during both HG1 and HG2 training stages. In the HG1 stage the robot learns three new higher-order action words (“GIVE”, “RECEIVE”, “KEEP”) by combining only basic action primitives; for example, the linguistic description for the higher-order word “KEEP” is “KEEP” [is] “GRASP” [and] “STOP”. In order to obtain the transfer of grounding from basic actions to higher-order words, the network calculates separately the output corresponding to the words contained in the description (“GRASP”, “STOP”) and stores it. Then, the network receives as input the higher-order word “KEEP” and as target the outputs previously stored.

The HG1 training stage runs for 30000 iterations and as shown in Fig. 5, after 5000 runs, the value of the error is smaller then 0.02.

The HG2 training stage runs for 30000 iterations as well (Fig. 6) and the value of the error is smaller then 0.02 after 5000 runs.

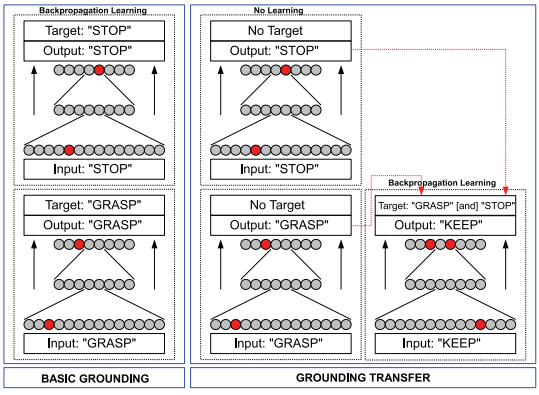


Fig. 4. Representation of the procedure that implements the grounding transfer mechanism

During the HG2 stage, the robot learns three higher-order behaviour (“ACCEPT”, “REJECT”, “PICK”) consisting of the combination of basic action primitives and higher-order action words (e.g. “ACCEPT” [is] “KEEP” [and] “SMILE” [and] “STOP”).

At the end of the training of the neural network all the actions primitives, higher-order words and higher-order behaviour were successfully learned.

Table I contains the simulation parameters used for training the feedforward NN and the Root Mean Square Error (RMSE) calculated at the end of each training stage.

TABLE I
SIMULATION PARAMETERS FOR THE TRAINING OF THE NETWORK AND RMSE

Training Stage	No. Iterations	Larn Rate	Momentum	RMSE
BG	30000	0.2	0.9	0.005840
HG1	30000	0.2	0.9	0.005620
HG2	30000	0.2	0.9	0.005042

B. Robot Simulation

For teaching the robot to perform actions associated to basic words, the Action Primitives library of the iCub soft-

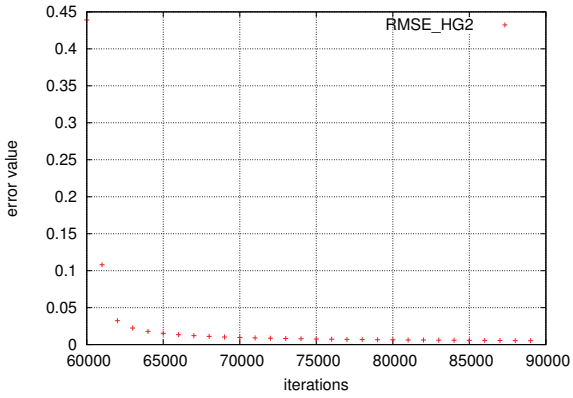


Fig. 6. Root Mean Square Error after the HG2 training stage

ware repository has been used. The library, that relies on the YARP Cartesian Interface [26], implements abstraction layers that allows the execution of higher level actions and more sophisticated tasks. The Action Primitives library provides a set of functions to perform basic action primitives and to combine them in order to obtain more complex behaviours. The functions contained in the library can move the arm of the robot to a specific position (e.g. MOVE_ARM_AWAY, MOVE_ARM_TOWARD), execute a predefined fingers sequence (OPEN_HAND, CLOSE_HAND) and wait for a specific time interval (STOP). For executing an action it is necessary to push the corresponding request item in the actions queue by using the function *pushAction (params)*; by using this function it is possible to insert in the action queue a sequence of simple actions to be executed.

A software module that relies on YARP and Action Primitives libraries has been developed in order to execute action primitives associated to verbal labels. Fig. 7 shows the execution of the basic action primitives "MOVE_ARM_AWAY" (Fig. 7(b)), "CLOSE_HAND" (Fig. 7(c)), "OPEN_HAND" (Fig. 7(d)) and "MOVE_ARM_TOWARD" (Fig. 7(e)).

The Action Primitives library relies on the YARP Cartesian Interface that allows the user to control upper limbs of the robot by defining a specific pose (position and orientation in axis-angle representation) for the end-effector [27]. In order to determine the joints configuration to move the robot arms to a desired position, a nonlinear optimization technique is used. The inverse kinematic is solved by using the IpOpt software package [28].

Through YARP, that is an open source framework for decoupling devices from software architecture, it is possible to exchange information between the user code and the simulated robot with its environment. The iCub Simulator and

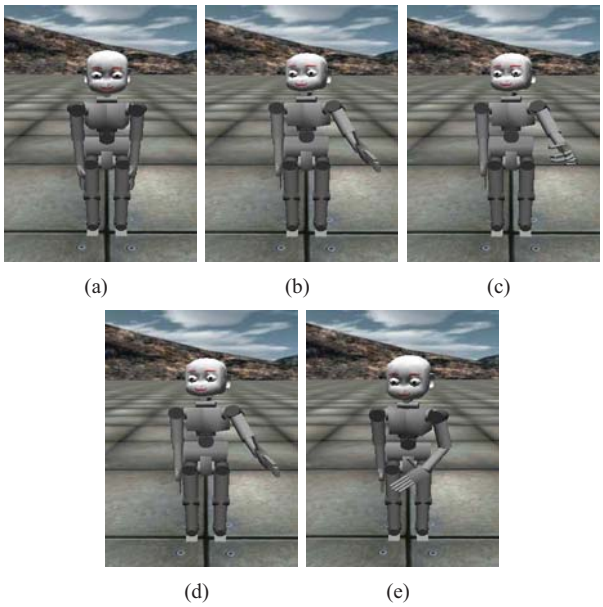


Fig. 7. Execution of basic action primitives on the iCub: initial position (a), "MOVE_ARM_AWAY" (b), "CLOSE_HAND" (c), "OPEN_HAND" (d) and "MOVE_ARM_TOWARD" (e)

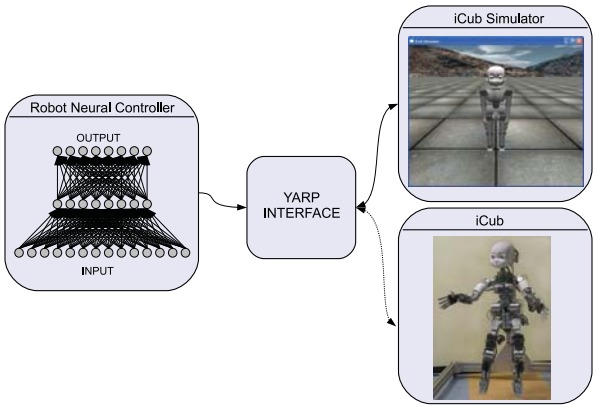


Fig. 8. Software architecture of the proposed model for the learning of words: Neural Network controller, YARP interface, iCub Simulator and real robotic platform

the real robot have the same software interface. Therefore, the adaptation of the simulated experiments to the physical robot (Fig. 8) will not require any particular modification of the code linking the neural network with the robot (though extra work is required to handle with visual input stream and motor performance).

C. Discussion

Building intelligent systems that can learn the meaning of abstract words is a challenging task for cognitive developmental robotics. In this paper an embodied neural network model that extends the symbol grounding mechanism to abstract words in cognitive agents has been presented. Simulation results show that the grounding of words that express general actions with a sensorimotor component can be obtained as a consequence of sensorimotor experiences; by combining words grounded in direct sensorimotor experience the simulated robot can acquire more abstract words.

The results of simulations demonstrate that new concepts can be learned through the mechanism of the grounding transfer; higher-order behaviours can be *indirectly-grounded* in basic action primitives *directly-grounded* in sensorimotor experience. In the field of research on cognitive robotics, the results achieved are very important because the acquisition of new concepts can be obtained as a combination of the previously grounded words. Through the grounding transfer mechanism a cognitive system can autonomously get more knowledge about the environment it interact with. The presented model can be scaling up in order to acquire more complex behaviours and linguistic abilities.

In the current work, the Higher-order Grounding mechanism, that implements the transfer of grounding, takes inspiration from the PSS theory proposed by Barsalou [1]. During the HG learning stages, the robot learns new higher-order words (e.g. "KEEP") by reactivating the internal "mental simulations" of the basic words contained in the linguistic description ("GRASP", "STOP") previously recorded. The proposed model can be extended to test other embodied

cognition theories of language learning as for example, the Action-sentence Compatibility Effect [29]. This model provides a useful tool for investigating and testing new hypotheses on embodied theory of language learning.

V. Conclusions

In the framework of cognitive robotics, the development of linguistic abilities is very important in order to achieve an effective interaction between human beings and cognitive autonomous agents. The current model can be extended in order to include in the robot both language comprehension and production capabilities; it can be useful for human-robot interaction and for easily programming the robot, for instance, through linguistic description provided by users through natural language. For human beings language skills are one of the most powerful tools for understanding situations and communicating with other agents in the environment. Therefore, the learning of words requires mental capabilities that involve different cognitive capacities working together [30]. Two of the most important approaches to language acquisition proposed in literature are Constructivist theories and Nativist theories [31]; according to the Constructivist theory, children gradually "construct" their grammar knowledge through experience [32] while for Nativist theory some aspects of the language are innate [33][34]. In the framework of the development of linguistic abilities, another distinction can be done between Domain Specific theories and Domain General theories; in the first case, cognitive processes are specialized and confined to handling just one specific type of information while according to the Domain General theories, cognitive processes are applicable to information in many different knowledge domain [31].

In this paper we focused on the learning of abstract language; abstract terms, on the contrary of concrete words, that refer to objects or events that can be perceived by senses, refer to things that are not tangible and that have no physical referents in the world. Then, while for concrete words is possible to attach labels to concepts formed on the basis of sensorimotor experience, is not possible to do the same for abstract words [35]; furthermore, abstract terms like *love* change meaning with time and circumstances, while concrete terms like *book* have a stable meaning. For all these reasons, obtaining a semantic representation of abstract words is changeling. In this study we focused on the learning of semi-abstract words that have motoric features; these kind of words refer to the way in which an object can be used or describe its motion (e.g. grasp, receive, etc.). In particular, in the experiments described above we taught the robot the meaning of words that express general actions and characterised by a sensorimotor component; this allowed us to build the meaning of words through direct sensorimotor experience. As an extension of the current work, the idea is to focus our attention to words with a major level of abstractness.

Furthermore, we would like to extend the presented model in order to include in the Neural Network controller the encoding of motor outputs. In the current work the NN controller receives linguistic inputs and gives as output a combination of the name of action primitives to be executed. We would like to implements a Neural Network model that controls both motor and linguistic behaviour of the robot. The idea is to have a neural controller that receives the linguistic input and calculates in output the motor response, that correspond to the force that should be applied to the robot's joints in order to perform a specific action. In this way the robot is able to simultaneously learn how to perform actions and their corresponding names; when a linguistic input is given to the robot, the neural controller activates the corresponding input node and calculates as output the force to be applied to the robot's joints.

References

[1] L.W. Barsalou, "Perceptual symbol systems," *Behavioral and brain sciences*, vol. 22, no. 4, pp. 577-660, Cambridge Univ Press, 1999.

[2] A. M. Glenberg, "What memory is for," *Behavioral and Brain Sciences*, vol. 20, pp. 1-55, 1997.

[3] G. Lakoff and M. Johnson, "Metaphors We Live By," *The University of Chicago Press*, 1980.

[4] G. Sandini, G. Metta, and D. Vernon, "The iCub cognitive humanoid robot: An open-system research platform for enactive cognition," *50 years of artificial intelligence*, pp. 358-369, Springer, 2007.

[5] A. Cangelosi and T. Riga, "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots," *Cognitive Science*, vol. 30, no. 4, pp. 673-689, 2006.

[6] A. Cangelosi, E. Hourdakis, and V. Tikhanoff, "Language acquisition and symbol grounding transfer with neural networks and cognitive robots," *Neural Networks, 2006. IJCNN'06. International Joint Conference on Neural Networks*, pp. 1576-1582, 2006.

[7] S. Harnad, "The symbol grounding problem," *Physica D: Nonlinear Phenomena*, vol. 42, no. 1-3, pp. 335-346, Elsevier, 1990.

[8] A. Cangelosi (in press), "Solutions and open challenges for the symbol grounding problem," *International Journal of Signs and Semiotic Systems*, (with commentaries).

[9] M.M. Louwerse, "Symbolic or embodied representations: A case for symbol interdependency," *Handbook of latent semantic analysis*, pp. 107-120, 2007.

[10] S. T. Dumais and T. K. Landauer, "A solution to Platos problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge," *Psychological review*, vol. 104, pp. 211-240, 1997.

[11] J. Searle, "Minds, brains, and programs," *Behavioral and Brain Sciences*, vol. 3, no. 3, pp. 417-457, 1980.

[12] A. Cangelosi, "Approaches to Grounding Symbols in Perceptual and Sensorimotor Categories," *Handbook of Categorization in Cognitive Science*, pp. 719-737, 2005.

[13] D.E. Rumelhart and J.L. McClelland, "Parallel distributed processing: explorations in the microstructure of cognition," , vol. 1, Foundations, MIT Press, Cambridge, Ma, 1986.

[14] C. Scorolli A.M. Borghi, "Sentence comprehension and action: Effector specific modulation of the motor system," *Brain research*, vol. 1130, pp. 119-124, 2007.

[15] L.W. Barsalou, "Grounded cognition," *Psychology*, vol. 59, no. 1, pp. 617-645, 2008.

[16] M. Andrews, G. Vigliocco and D. Vinson, "Integrating experiential and distributional data to learn semantic representations," *Psychological review*, vol. 116, no. 3, pp. 463-498, 2009.

[17] G. Lakoff, "Women, fire, and dangerous things: What categories reveal about the mind," *The University of Chicago Press*, 1987.

[18] A. Cangelosi, A. Greco and S.Harnad, "From Robotic Toil to Symbolic Theft: Grounding Transfer from Entry-Level to Higher-Level Categories," *Connection Science*, vol. 12, no. 2, pp. 143-162, 2000.

[19] A. Greco, T. Riga, and A. Cangelosi, "The acquisition of new categories through grounded symbols: An extended connectionist model," *Artificial Neural Networks and Neural Information Processing - ICANN/ICONIP 2003*, Berlin, pp. 773-770, 2003.

[20] T. Riga, A. Cangelosi and A. Greco, "Symbol grounding transfer with hybrid self-organizing/supervised neural networks," *IEEE International Joint Conference on Neural Networks - Proceedings*, vol. 4, pp. 2865-2869, 2004.

[21] A. Cangelosi, "Symbol Grounding in Connectionist and Adaptive Agent Models," *New Computational Paradigms, Lecture Notes in Computer Science*, vol. 3526, pp. 69-74, Springer Berlin / Heidelberg, 2005.

[22] A. Cangelosi, "Modeling Language, Cognition and Action: From connectionist simulations to embodied neural cognitive systems," *Modelling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*, vol. 16, pp. 3-14, Singapore: World Scientific, 2005.

[23] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The iCub humanoid robot: an open platform for research in embodied cognition," *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pp. 50-56, 2008.

[24] V. Tikhanoff, A. Cangelosi, P. Fitzpatrick, G. Metta, L. Natale, and F. Nori, "An open-source simulator for cognitive robotics research: The prototype of the icub humanoid robot simulator," *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pp. 57-61, 2008.

[25] V. Tikhanoff, A. Cangelosi and G. Metta (in press), "Language understanding in humanoid robots: Simulation experiments with iCub platform," *Proceedings of 2009 International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS09)*, St. Louis MA, 2009.

[26] G. Metta, P. Fitzpatrick and L. Natale, "Yarp: yet another robot platform," *International Journal of Advanced Robotic Systems*, vol. 3, no. 1, pp. 43-48, 2006.

[27] U. Pattacini, F. Nori, L. Natale, G. Metta and G. Sandini, "An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots," *2010 IEEE/RSJ international conference on intelligent robots and systems*, pp. 1668 - 1674, Taipei, 2010.

[28] A. Wächter and L.T. Biegler, "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming," *Mathematical Programming*, vol. 106, no. 1, pp. 25-57, 2006.

[29] A. M. Glenberg and M. P. Kaschak, "Grounding language in action," *Psychonomic Bulletin & Review*, vol. 9, no. 3, pp. 558-565, 2002.

[30] P. Bloom, "How children learn the meanings of words," *The MIT Press*, 2002.

[31] M.D. Barrett, "The development of language," *Psychology Pr*, 1999.

[32] M. Tomasello, "Constructing a language: A usage-based theory of language acquisition," *Harvard University Press*, 2003.

[33] N. Chomsky, "Knowledge of language: Its nature, origin, and use," *Praeger Publishers*, 1986.

[34] S. Pinker, "Learnability and cognition: The acquisition of argument structure," *MIT press Cambridge, MA*, 1989.

[35] A.M. Borghi and F. Cimatti, "Words as tools and the problem of abstract words meanings," *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, vol. 31, pp. 2304-2309, 2009.